

## COMPUTACIÓN E INFORMÁTICA

*Recibido 11 Jul 2016 Aceptado 14 Oct 2016*

*ReCIBE, Año 5 No. 3, Noviembre 2016*

# Evidencia Empírica de la Minería de Procesos en la Implantación de CMMI-DEV

## Empiric Evidence of Process Mining in CMMI-DEV Implementation

Paola E. Velazquez-Solis<sup>1</sup>  
paola.velazquez@uabc.edu.mx

Brenda L. Flores-Rios<sup>1</sup>  
brenda.flores@uabc.edu.mx

María Angélica Astorga-Vargas<sup>1</sup>  
angelicaastorga@uabc.edu.mx

Jorge E. Ibarra Esquer<sup>1</sup>  
jorge.ibarra@uabc.edu.mx

Félix Fernando González Navarro<sup>1</sup>  
feranando.gonzalez@uabc.edu.mx

Francisco J. Pino<sup>2</sup>  
fjpino@unicauca.edu.co

<sup>1</sup> Universidad Autónoma de Baja California Campus Mexicali, Mexicali, B.C., México

<sup>2</sup>Universidad del Cauca, Colombia

**Resumen:** La minería de procesos tiene como objetivo descubrir, monitorear y mejorar procesos a través del análisis de los diversos registros de eventos generados por los procesos de la organización. El objetivo de este trabajo es presentar la evidencia empírica de la inclusión estratégica de la disciplina de minería de procesos en proyectos de mejora de procesos de software implementados con CMMI. En el proceso de mapeo sistemático de la revisión de la literatura, se establecieron cuatro categorías para clasificar los hallazgos encontrados (Fundamentos teóricos, propuestas, herramientas y sistemas de información y algoritmos) para presentar los estudios que cumplen con el objetivo. Se concluye que la interdisciplinariedad de la minería de procesos con un modelo de referencia de procesos como CMMI-DEV apoya la implementación y evaluación de las áreas de procesos, al aplicar técnicas y algoritmos de minería de procesos que faciliten la exploración y explotación de los registros de eventos relacionados a la ejecución de las actividades almacenados en un repositorio.

**Palabras clave:** Minería de Procesos, Mejora de Procesos de Software, Registro de Eventos.

**Abstract:** The discipline of process mining aims to discover, monitor and improve processes through the analysis of the various event logs generated by the organization's processes. In such a context, the goal of this paper is to present empirical evidence of the strategic inclusion this discipline into software process improvement projects based on CMMI. In the process of mapping the results of a systematic literature review, four categories were established to classify the findings (Theoretical Foundations, proposals, tools and information systems and algorithms) and to present studies that achieve the objective established. We conclude that the interdisciplinary collaboration of process mining with a reference process model such as CMMI-DEV supports the implementation and evaluation of process areas, applying techniques and algorithms of process mining that facilitate the exploration and exploitation of event logs related to the implementation of activities stored in a repository.

**Keywords:** Process Mining, Software Process Improvement, Event Log.

# 1. Introducción

En un proyecto de mejora de procesos de software (SPI por sus siglas de Software Process Improvement), en la medida que se genera la evidencia de la ejecución de las actividades o prácticas, requeridas por el modelo de referencia de procesos o ciclo de mejora, se incrementa la cantidad de datos e información por cada proceso, provocando algunos problemas en la escalabilidad o control de la información almacenada en el repositorio de evidencias de la organización. Por lo general, se utilizan como repositorios las bases de datos, bases de conocimiento, intranet o sistemas de información donde el tamaño se basa en criterios de capacidad y eficiencia de almacenamiento y no en su uso posterior, mucho menos en la exploración, explotación y análisis de los registros. Es así, como el limitado acceso a dicha información impide el enriquecimiento de criterios para la toma de decisiones necesarias en el descubrimiento de nuevas reglas de negocio, mejora continua, monitoreo o un control de los procesos ya establecidos (Orellana García et al., 2015).

Al revisar los diferentes estándares y modelos de procesos de software, como CMMI, ISO/IEC 15504, ISO/IEC 12207:2004, NMX-I-059-NYCE-2011 y COMPETISOFT (Pino et al., 2006), se detectó que no plantean técnicas, herramientas o algoritmos para extraer, de manera adecuada, la información o el conocimiento existente en sus registros (registro de eventos). Específicamente para dar evidencia de la implantación del modelo de referencia CMMI, se utiliza la revisión de la ejecución de las prácticas y artefactos almacenados en los repositorios, además de entrevistas dirigidas a los roles involucrados. De esta forma, los registros están compuestos por instancias del área de procesos y cada instancia contiene la información relacionada con la ejecución de la práctica como nombre de la actividad, rol responsable y tiempo entre otros elementos. Siendo los registros de eventos una fuente de datos para el análisis de la madurez organizacional. En la representación escalonada se establecen cinco

niveles de madurez (del 1 al 5) en función de un grupo de áreas de procesos (Chrissis, 2011). Desde el nivel 2 (Gestionado) se garantiza que las áreas de procesos se planifican y ejecutan de acuerdo con las políticas; los proyectos de software se conforman con personal calificado que dispone de recursos adecuados para producir resultados controlados (Chrissis, 2011). Al revisar el Maturity profile del 2015 de CMMI v1.3, el mayor número de evaluaciones corresponde a organizaciones con nivel 3 de madurez (Definido) (CMMI Institute, 2015), considerándolo el nivel que requiere más esfuerzo y conocimiento técnico, en función al mayor número de áreas de procesos asociados a las categorías que se evalúan dentro de éste.

Bajo este contexto, en este documento se expone la importancia de implementar técnicas o algoritmos que permitan ser adoptados o aplicados por las organizaciones de desarrollo de software para conocer el estado actual de sus procesos y/o diagnosticar el nivel de capacidad de los mismos en un tiempo determinado. Una estrategia es apoyarse de la disciplina de minería de procesos, la cual tiene como objetivo descubrir, monitorear y mejorar procesos a través de la extracción y exploración del conocimiento contenidos en los registros de eventos de los sistemas de información (van der Aalst, 2011; Orellana García et al., 2015). Los registros de eventos corresponden al histórico de ejecución de los procesos, donde se encuentran las instancias o casos del proceso (Orellana García et al., 2016). Cuando una organización cuente con un proyecto SPI y genere un volumen de información, la minería de procesos brindaría las técnicas, algoritmos o herramientas con el objetivo de llevar a cabo la exploración de los registros de eventos (Yzquierdo et al., 2013) con una mayor cobertura y confiabilidad, mientras se mantienen menores costos y esfuerzos (Valle et al., 2014). De esta forma, los procesos implementados, la información y el conocimiento explorado y descubierto, estarían perfectamente alineados a los requerimientos relacionados con el cumplimiento de normas, eficiencia y servicios del cliente (van der Aalst et al., 2012).

Este documento proporciona un mapeo sistemático como una evidencia empírica dirigida a la comunidad de Ingeniería de software e investigadores de Inteligencia Artificial presentando la interrelación de la minería de procesos en proyectos SPI basados en CMMI. La técnica de mapeo sistemático define un proceso y un informe que permite categorizar los resultados que han sido publicados hasta el momento en un área determinada (Petersen et al., 2008). Los resultados son analizados según su aportación metodológica, casos de estudio y herramientas empleadas para la exploración de los registros de eventos. Así mismo, el documento se estructura de la siguiente manera: en la sección 2 se aborda el marco conceptual con información base sobre la minería de datos, minería de procesos y mejora de procesos de software. La sección 3 describe la metodología de utilizada en el mapeo sistemático de los trabajos de minería de procesos en proyectos SPI basados en CMMI; la sección 4 presenta los resultados, ventajas y desventajas identificados en cada uno de ellos. Por último, se presentan las conclusiones y trabajo futuro.

## 2. Marco Conceptual

La minería de procesos es una disciplina relativamente joven que las organizaciones están incorporando a sus proyectos y aplicaciones con el objetivo de mejorar sus procesos de negocio (Orellana García et al., 2015), por lo que es importante distinguir la diferencia entre la minería de datos y la minería de procesos. Mientras la primera es utilizada para encontrar patrones en grandes volúmenes de datos, en una relación de patrones y datos; la minería de procesos encuentra las relaciones de los procesos de la organización dentro de los datos, explorando la información de los registros; es decir, se sitúa entre la minería de datos y el análisis de procesos, implementando Business Intelligence (van der Aalst, 2011) y Business Process Management, que tiene como objetivo el descubrimiento de modelos basados en el registro de eventos disponibles (Gupta, 2014).

En la siguiente sección, se presenta el estado del arte de registros de eventos, minería de procesos y mejora de procesos de software basados en CMMI-DEV.

## 2.1 Registro de Eventos

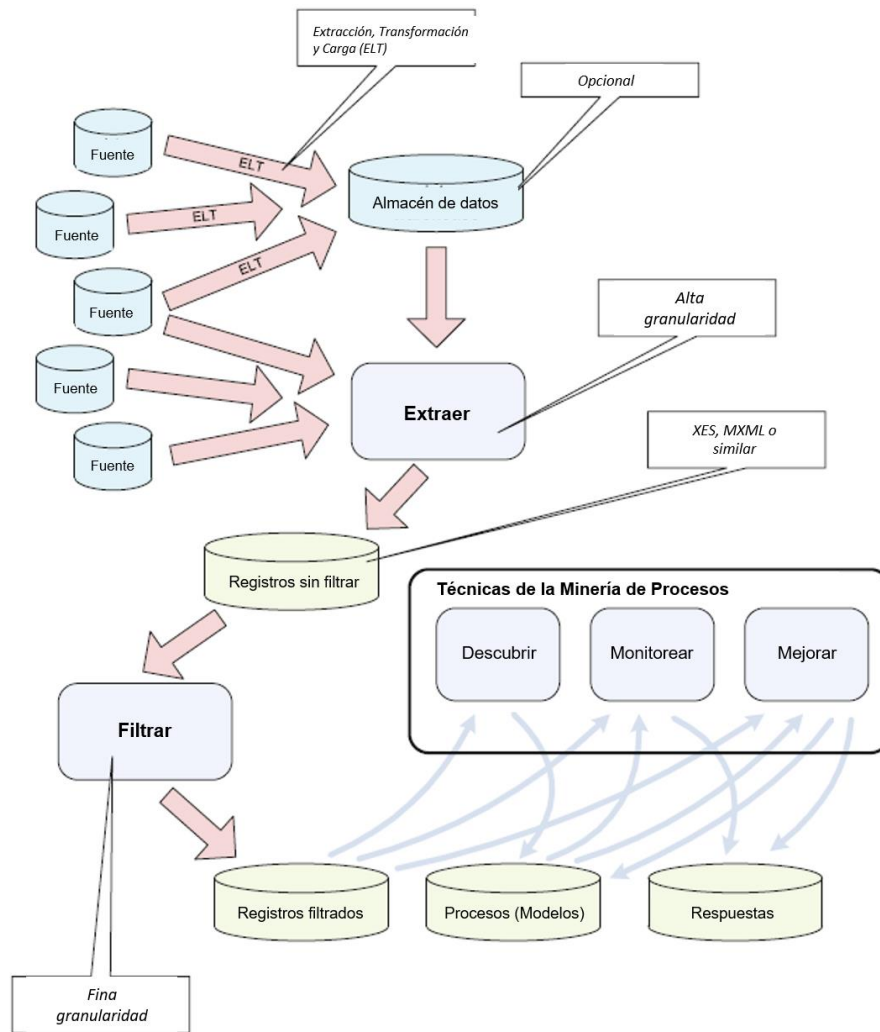
Los registros de eventos tienen como objetivo extraer información a partir de los datos almacenados en los registros de trazas de los sistemas de información, en busca de errores inconsistencies, vulnerabilidades y variabilidad en los procesos que se ejecutan (Orellana García et al., 2016). El propósito es analizar los datos de eventos desde la perspectiva orientada a procesos.

La Figura 1 muestra el flujo de trabajo de los registros de eventos, donde el punto de partida son los datos en bruto ocultos en todo tipo de fuentes de datos. Una fuente de datos puede ser un archivo plano simple, una hoja de cálculo de Excel, un registro de transacciones, o una tabla de base de datos (van der Aalst, 2011). Los datos de los registros de eventos, con mayor frecuencia, se dispersan típicamente a través de diferentes fuentes, requiriendo grandes esfuerzos para recopilar los datos pertinentes. Los eventos también pueden capturarse buscando en los intercambios de mensajes, el registro de lectura y escritura y también las fuentes de datos que son estructurados por metadatos. Desafortunadamente, en muchas situaciones, los datos son datos no estructurados (Verdonk et al., 2010).

En la misma figura, se observan flujos de salida desde las fuentes de datos que indican la leyenda ETL refiriéndose a la extracción, transformación y carga (ETL). ETL se utiliza para describir el proceso que consiste en: a) extracción de datos de fuentes externas, b) transformación de las necesidades operacionales, y c) la carga en el sistema de destino en un almacén de datos o base de datos relacional. Un almacén de datos es un repositorio único lógico de los datos transaccionales y operativos de una organización. El almacén de datos no produce datos. El objetivo es unificar la información de tal manera que se puede utilizar para la presentación de informes, análisis, pronóstico, etc. Dependiendo de las preguntas y puntos de vista elegidos, diferentes registros de eventos

pueden ser extraídos del mismo conjunto de datos, basándose en el registro filtrado, dependiendo de los diferentes tipos de minería de procesos que se deseen aplicar: descubrimiento, de conformidad y mejora (van der Aalst, 2011). Estas tres técnicas se explican más adelante.

Es importante señalar que aunque la Figura 1 no refleja la naturaleza iterativa de todo el proceso, es probable que los resultados de la minería de procesos desencadenen nuevas preguntas que a su vez, conduzcan a la exploración de nuevas fuentes de datos y extracciones de datos más detallados. Por lo general, se necesitan varias iteraciones de la extracción, filtrado y fases de la producción. Algunos sistemas de información PAIS (por las siglas de Processs Aware Information Systems) del tipo workflow, BPMS, ERP, CRM, entre otros (Pérez Alfonso et al., 2012) permiten generar automáticamente los registros de eventos, descubrir modelos de procesos considerando distintos objetivos o perspectivas como el control de flujo, organizacional, o temporal por medio de diversas representaciones (Pizarro Álvarez, 2013). De esta forma, los registros de eventos almacenan la información de los eventos asociados a las actividades de los procesos (Pérez Alfonso et al., 2012).



**Figura 1:** Diagrama de extracción de datos para un registro de eventos.  
Traducida de (Van der Aalst, 2011).

## 2.2 Minería de Procesos

El concepto de minería de procesos ha evolucionado a través de los años. En un inicio se definió como una técnica que utiliza los registros de flujo de trabajo para descubrir el proceso que en realidad está ejecutándose (Weijters & van der Aalst, 2001). Posteriormente, con la creación de nuevos algoritmos para el análisis de los registros de eventos, la minería de procesos se definió como un conjunto de técnicas y procedimientos que trataban de reconstruir de forma



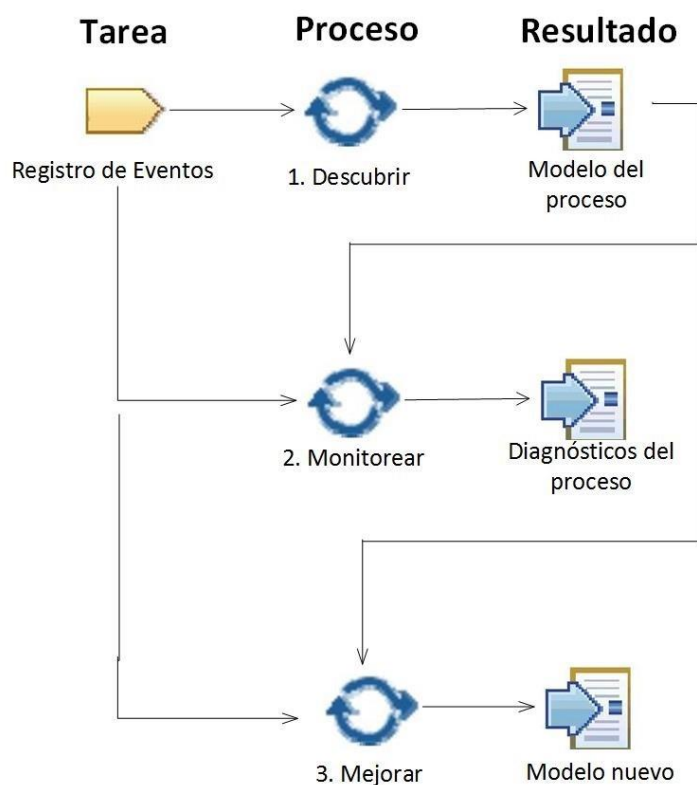
automática el modelo observado en los registros de eventos (van der Aalst et al., 2004). Antes de convertirse formalmente como una disciplina ante la IEEE en el 2011, la minería de procesos se refería a un grupo de técnicas con el objetivo de descubrir, mejorar y monitorear los procesos reales a través de la extracción del conocimiento de un conjunto de ejecuciones reales de procesos (Nakatumba, 2009). Hoy en día, los registros de eventos ya están automatizados en algunos sistemas de información (van der Aalst, 2011; Orellana García et al., 2015).

El objetivo de la minería de procesos es responder a preguntas de la perspectiva operacional de los procesos. Por ejemplo: ¿Qué sucedió realmente en el pasado?, ¿Por qué sucedió?, ¿Qué es probable que suceda en el futuro?, ¿Cuándo y por qué las organizaciones y personas se desvían?, ¿Cómo controlar mejor un proceso?, ¿Cómo rediseñar un proceso para mejorar su rendimiento? Los autores (van der Aalst et al., 2011), especifican que la minería de procesos brinda las técnicas que ayudan a las organizaciones a:

1. Descubrir el modelo de ejecución real del proceso. Los algoritmos de minería de procesos cuando son aplicados al análisis de los registros de eventos, permiten generar el modelo real de un proceso, el cual es expresado en una red de Petri o en la notación BPMN (por sus siglas de Business Process Modeling Notation). Dicho modelo representa la situación real y no se basa en la documentación de cómo se ejecuta el proceso o en la percepción de los involucrados en el mismo.
2. Determinar si el proceso cumple con la reglamentación y los procedimientos documentados. Comparar el modelo obtenido con los procesos documentados para determinar el cumplimiento de estándares, protocolos, reglamentación y políticas de ejecución de un proceso.
3. Analizar la interacción del personal que ejecuta el proceso. Brinda la posibilidad de construir redes sociales del proceso para analizar la interacción y descubrir posibles bucles que demoran el proceso.
4. Descubrir cuellos de botella. Especificar actividades tipo cuello de botella.
5. Determinar la relación entre las variables de un caso, por medio de la aplicación de técnicas de clasificación.

En la Figura 1 se indicó que las técnicas de minería de procesos asumen que es posible registrar eventos de manera secuencial, de tal manera que cada uno de los eventos se refiere a una actividad; es decir, un paso bien definido en algún

proceso y a su vez éste se relacione a un caso particular, como una instancia del proceso. En la Figura 2 se presenta un diagrama para estructurar el flujo de actividades para la implementación de las técnicas de: Descubrir, Monitorear y Mejorar.



**Figura 2:** Técnicas de Minería de Procesos. Traducida y adaptada de (van der Aalst, 2011).

En la Figura 2 se muestra cómo la primera técnica de minería de procesos está basada en el descubrimiento. Esta es una de las técnicas más destacadas (van der Aalst, 2011), la cual consiste en tomar como entrada un registro de eventos y generar un modelo del proceso sin usar ninguna información a-priori. La segunda técnica de minería de procesos se basa en el monitoreo y la conformidad. Es aquí donde se compara un modelo que ya existe contra un registro de eventos del mismo proceso (van der Aalst et al., 2012). La verificación

de conformidad puede ser usada para asegurarse si la realidad del diagnóstico obtenido coincide con los registros de eventos y el modelo del proceso, y viceversa. Mientras que la segunda técnica busca medir dicha conformidad, el tercer tipo de técnica de minería de procesos busca mejorar o extender un modelo ya existente (Wang et al., 2013). El objetivo es extender o mejorar un modelo existente utilizando la información acerca de la ejecución de un proceso real almacenada en algún registro de eventos. La Figura 2 presenta una relación entre la información requerida para la implementación de las tres técnicas de minería de procesos, así como la salida de una puede alimentar la entrada de otra.

## 2.3 Mejora de Procesos de Software en CMMI

Un programa SPI permite orientar a las organizaciones en la implementación, evaluación y mejora de sus procesos (Niazi et al., 2005). El principal enfoque de SPI está basado en el modelo, el cual compara los procesos de la organización con un modelo de referencia de procesos y utiliza una evaluación SPI para identificar debilidades y posibles opciones de mejora. Un ejemplo del enfoque SPI basado en modelo es CMMI (Vasconcellos et al., 2017). SPI basado en CMMI significa una estrategia formal para las organizaciones de software en lograr una implementación exitosa del modelo (Coleman & O'Connor, 2008) y aportar importantes beneficios para ellas, como el incremento de la satisfacción del cliente, la eficiencia del proceso de desarrollo y la definición y cumplimiento de los objetivos de calidad (Durón del Villar & Muñoz-Mata, 2013).

Para la implementación de un programa SPI las etapas a realizar son: 1) Diagnóstico de procesos actual de la organización en función de los requisitos normativos a través de la evaluación y la correspondencia de los procesos de la organización con el modelo de referencia; 2) Establecimiento de la estrategia de implantación con base en las necesidades de mejora identificadas en la evaluación diagnóstico y la definición del número de ciclos de mejora para cumplir con el nivel de madurez de la organización definido en su alcance; y 3)

Evaluación de los resultados obtenidos y mejora de los procesos implantados (Flores Rios et al., 2014). En este caso, el método de evaluación de CMMI es conocido como SCAMPI (por las siglas de Standard CMMI Appraisal Method for Process Improvement).

Existen organizaciones que aún enfrentan desafíos en la implementación de SPI basados en CMMI (Niazi, 2015), entre ellos la necesidad de contar con herramientas de software como un soporte valioso para el equipo de implementación, evaluación o equipo de mejora permitiéndoles recopilar, registrar clasificar y evaluar la evidencia sobre todo cuando el volumen y complejidad de la información son significativos (NYCE, 2010; Astorga et al., 2014). Otro desafío para las organizaciones es el mantener actualizados sus procesos de acuerdo a los flujos de trabajo reales con los que operan (van der Aalst, 2011).

Desde la perspectiva de la minería de procesos, los flujos de trabajo están incluidos implícitamente en cada registro de evento (actividad) asociado a cada uno de los procesos de la organización, permitiendo de esta manera realizar el descubrimiento, la mejora y el monitoreo de los procesos a través de la aplicación de herramientas de software para el análisis de datos dependiendo del ciclo de mejora y/o el nivel de madurez al que desee llegar la organización de software.

### 3. Metodología

Con el objetivo de proporcionar evidencia empírica que muestre los trabajos publicados hasta el momento sobre la incidencia de la disciplina de la minería de procesos en proyectos SPI basados en CMMI, en esta sección se describen las fases del mapeo sistemático como etapa previa a la revisión sistemática de la Literatura propuesta por Kitchenham (2007). La técnica de mapeo sistemático permite categorizar los resultados que se publican en una determinada área de interés (Martínez et al., 2011). El modelo de referencia de procesos CMMI se ha

sostenido por más de dos décadas en la industria de software, tomándolo como referencia para la generación de otros estándares o modelos como NMX-059-NYCE-2011 y COMPETISOF (Oktaba, 2008). Así mismo, se ha probado que CMMI puede ser implementado no sólo en grandes organizaciones sino también por Pequeñas y Medianas Empresas (PyMEs) (Oktaba, 2008; Lester et al., 2010) otorgándoles una ventaja competitiva en el mercado internacional. México, desde el 2002 ha venido impulsando la estrategia de implantación de CMMI en PyMES, contando actualmente con 180 centros de desarrollo certificados en CMMI-DEV (Secretaría de Economía, 2016).

### 3.1 Preguntas de investigación

Esta fase se centra en la identificación de los hallazgos en la adopción de minería de procesos como soporte a proyectos SPI. Se han definido las siguientes preguntas de investigación:

- a) ¿Existen publicaciones sobre la implantación de proyectos SPI usando la minería de procesos? y
- b) ¿Cuáles son los principales hallazgos de la adopción de minería de procesos en proyectos SPI basados en CMMI?
- c) ¿Qué tipo de calidad de estudios se han publicado para reportar los resultados de la adopción de la minería de procesos en proyectos SPI basados en CMMI?

La primera pregunta se refiere a la identificación de evidencia sobre la minería de procesos en la implantación de un proyecto SPI independientemente del modelo para llevar a cabo la mejora de procesos o un modelo de referencia de procesos o los métodos de evaluación de procesos a seguir. Las siguientes preguntas hacen hincapié en la identificación de hallazgos, las ventajas y desventajas y tipos de calidad de estudios en la adopción de la minería de procesos específicamente con el modelo CMMI.

El objetivo es buscar referencias y publicaciones sobre la minería de procesos, sus técnicas y algoritmos utilizados según el tipo de estudio o clasificación

(Fundamentos teóricos, Propuestas, Algoritmos, Sistemas de información) que han apoyado a un proyecto SPI basado en CMMI. Esta búsqueda se realiza con el apoyo de principales fuentes de información con enfoque técnico-científico, bajo los términos o cadenas de búsqueda claves de minería de procesos y procesos de software, durante el periodo 2005-2015.

## 3.2 Ejecución de la búsqueda

La búsqueda de literatura se realizó en cuatro fuentes electrónicas: Google Scholar, Researchgate, Scopus y Springer. El mapeo sistemático permite identificar, evaluar, interpretar y sintetizar todas las investigaciones existentes y relevantes en un tema de interés particular (Petersen et al., 2008). Las publicaciones más recientes dentro del período de búsqueda seleccionado en revistas especializadas, actas de conferencias e Internet (Kitchenham, 2007). Se utilizó una estructura de búsqueda, basado en conectores de palabras clave y lógica, obteniendo la siguiente cadena de búsqueda: "Process Mining" AND "Software Process Improvement".

En la primera búsqueda, se obtuvieron 18,641 documentos los cuales especifican la definición, los objetivos y propósitos de la minería de procesos. Durante 2001 a 2005, los autores (Weijters & van der Aalst, 2001; van der Aalst et al., 2007; Weijters, 2003), dieron las primeras propuestas sobre la minería de procesos como una herramienta y/o técnica para analizar cualquier proceso en general. Se encontraron casos de estudio sobre la implementación de la minería de procesos en procesos hospitalarios, sistemas industriales y dependencias gubernamentales, reportando que: a) en los sistemas hospitalarios se permitió monitorear la utilización de los recursos, como los implementos quirúrgicos, detectar funcionamientos anómalos, como la desviación, el exceso en el tiempo de estancia de un producto o insumo en el almacén o su ausencia, las variaciones entre la realidad y lo prescrito (Orellana García et al., 2015); b) en los sistemas industriales ha permitido analizar la tramitación de las facturas enviadas por los distintos subcontratistas y proveedores (Verdonk et al., 2010);

c) en auditorias en instituciones gubernamentales se logró encontrar violaciones de las normas específicas impuestas por la ley, indicar posibles fraudes, negligencias, riesgos, o ineficiencias (van der Aalst, 2011).

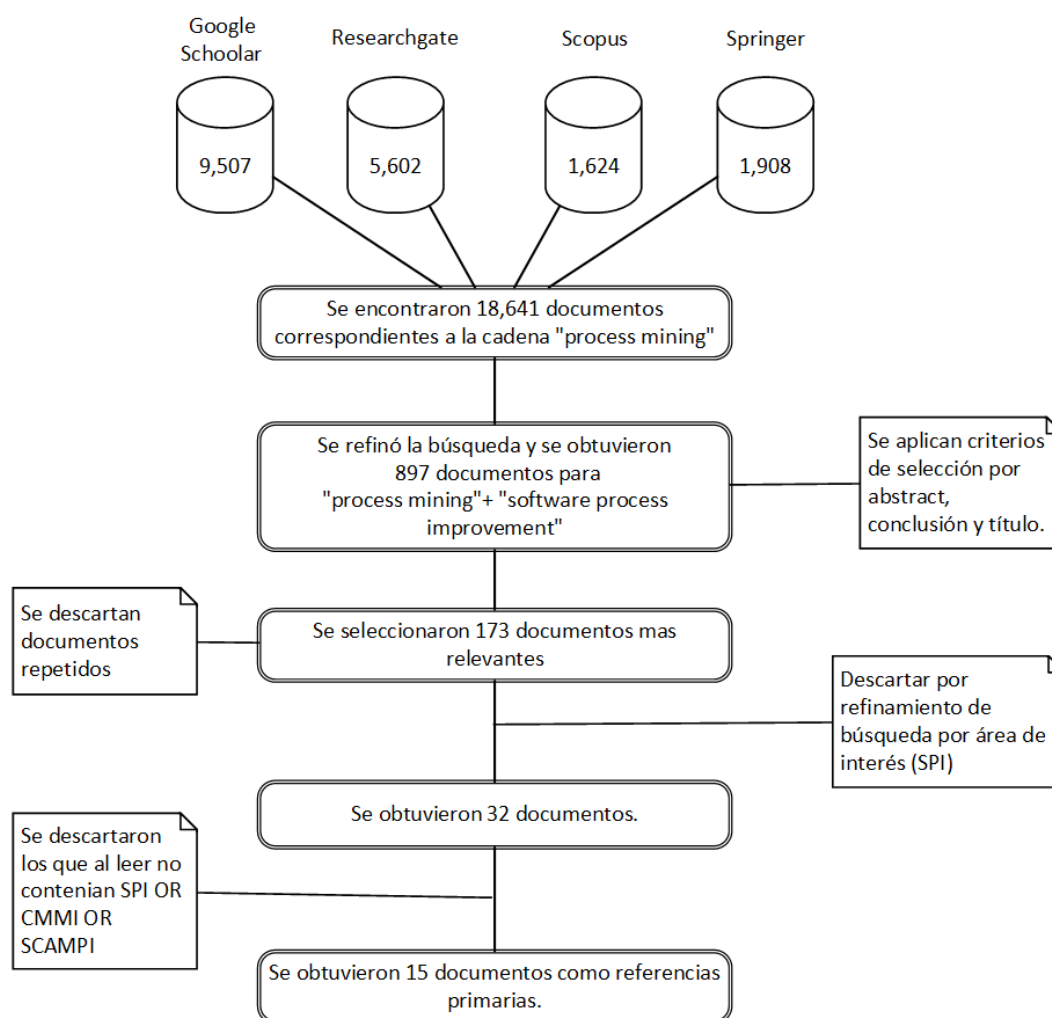
Para segmentar los documentos que asociaran a la minería de procesos con SPI se refinó la cadena de búsqueda “process mining” AND “software process improvement”, obteniendo como resultado 897 documentos. La cadena de búsqueda se adaptó a cada biblioteca electrónica y se restringieron las coincidencias en título, resumen y palabras clave. En el transcurso del 2015, el mayor número de evaluaciones de CMMI-DEV v1.3 corresponden con el nivel de madurez 3 (Definido), lo que indicaría ser el nivel que requiere más esfuerzo, tomando en cuenta el mayor número de áreas de procesos asociadas que se evalúan, sin considerar los niveles inferiores (CMMI Institute, 2015). Se consideraron las publicaciones que incluían las palabras clave: "software process improvement" OR "CMMI-DEV" OR "SCAMPI".

### 3.3 Filtrado de estudios

Con el fin de identificar los estudios más relevantes, se utilizó el siguiente procedimiento:

1. Selección de las palabras clave de la investigación;
2. Observación en cada biblioteca digital, tratando con las palabras clave de la investigación basadas en los criterios de inclusión y exclusión. Para la selección de los estudios primarios se definieron los siguientes criterios de inclusión: documentos, informes técnicos, artículos técnicos-científicos que presenten la inserción de la minería de procesos en SPI. Los trabajos seleccionados tuvieron que ser publicados entre 2005 y 2015 en revistas y conferencias, escrito en idioma Inglés. Se excluyeron aquellos resultados donde sólo se tenía acceso al resumen.
3. Análisis de cada documento a través del título y el resumen;
4. Descarga de documentos que cubren los criterios de búsqueda;
5. Análisis de la introducción y las conclusiones a conciencia, y teniendo una visión rápida del resto de las secciones en un documento. Se aplicó este proceso en cada biblioteca electrónica.

Con base en criterios de resumen y título, se aplicó un nuevo filtrado de los estudios reduciendo el número a 173 documentos. Se descartaron los que no consideraban los términos de SPI, CMMI y SCAMPI; así como los documentos que contenían información repetida. Finalmente, se obtuvo un total de 15 documentos (Figura 3). En 11 documentos de ellos, se plantea la interdisciplinariedad de la minería de procesos aplicada a proyectos SPI. En los 4 restantes, se enfocan específicamente al modelo de referencia de procesos CMMI utilizando la minería de procesos como una herramienta para la exploración de la información. Este mapeo sistemático permitió observar que del conjunto seleccionado de 15 documentos finales el 26.66% está relacionado con el objetivo de esta investigación.





**Figura 3:** *Proceso de ejecución de la búsqueda y filtrado de resultados.*  
*Elaboración propia.*

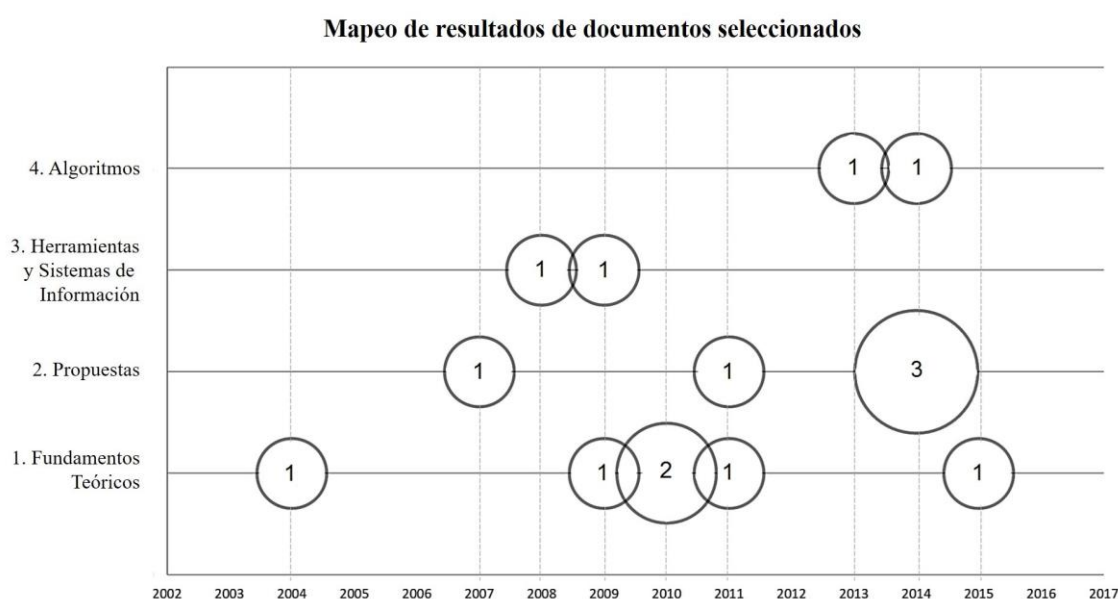
### 3.4. Esquema de clasificación

La Figura 3 muestra los 15 documentos relevantes seleccionados los cuales fueron clasificados en: 1) Fundamentos teóricos, 2) Propuestas, 3) Herramientas y Sistemas de Información y 4) Algoritmos. Partiendo de que desde hace 4 años la minería de procesos es una disciplina formal, la categoría 1) Fundamentos teóricos se relaciona con el marco conceptual base para los investigadores y profesionales. La categoría 2) Propuestas contribuye a responder la pregunta de investigación 2 presentada en la sección 3.1. Esta categoría muestra los principales hallazgos de la adopción de la minería de procesos en proyectos SPI, específicamente basados en CMMI. Los registros de eventos se han explotado con ayuda de herramientas disponibles y se han desarrollado varios sistemas académicos y comerciales (van der Aalst & Van Dongen, 2013) por tal motivo, la categoría 3 se relaciona a la identificación y caracterización de herramientas y sistemas de información. Por último, la categoría 4 se refiere a los algoritmos o tipo de técnicas de minería de procesos que apoyan en gran medida a la resolución de la problemática en una organización de desarrollo de software en la exploración y explotación del conocimiento y registro de eventos brindando así la evidencia de la ejecución de actividades en un proyecto SPI implementando CMMI-DEV.

### 3.5 Mapeo

El último paso del mapeo sistemático consiste en generar un modelo (mapa visual) de las dimensiones. La Figura 4 muestra el diagrama de burbuja que visualiza la frecuencia de documentos por años de publicación y la intersección de cada uno de los documentos con su clasificación asignada. Esto significa que el tamaño de una burbuja es proporcional al número de documentos reportados en el año y en relación a una única categoría. Por ejemplo, la categoría 2)

Propuestas visualiza la burbuja con mayor tamaño con un total de 3 documentos en el 2014. Otra vista es la frecuencia de documentos publicados por años, donde la categoría 1) Fundamentos teóricos cuenta con 6 documentos en un periodo del 2004 al 2015, concentrándose más en el periodo del 2009 al 2011. De este modo, las dos primeras categorías reflejan, de manera visual, un mayor esfuerzo en establecer los fundamentos teóricos y las propuestas sobre la minería de procesos en proyectos SPI implementando el modelo de referencia CMMI-DEV; a diferencia de las categorías 3 y 4 con las menores frecuencias. De igual forma, en la Figura 4 se observa que a partir del 2007 la frecuencia de los documentos va en aumento por año de publicación, siendo el año de 2014 el de mayor número con 4 documentos relacionados a las categorías de 2) Propuestas y 4) Algoritmos.



**Figura 4:** Visualización del mapeo sistemático en diagrama de burbuja.  
Elaboración propia.

## 4. Análisis y Discusión

Durante 15 años, los aportes más completos para los principios de la minería de procesos fueron hechos por Wil M. P. van der Aalst (2001, 2003, 2004, 2007, 2011, 2012, 2013). En la Tabla I se presenta una relación de definiciones del término de minería de procesos, sugeridas por diversos autores, las cuales han ido en evolución al incluir los análisis de los registros de eventos y las técnicas, algoritmos y herramientas basadas en principios de minería de datos, Business Intelligence y Business Process Management. Actualmente, la minería de procesos es reconocida como una disciplina ante la IEEE (van der Aalst et al., 2012).

**Tabla 1.** Relación de propuestas para la definición de minería de procesos

Año	Autor	Definición
2001	van der Aalst	Técnica que utiliza los registros de flujo de trabajo para descubrir el proceso que en realidad está ejecutándose (van der Aalst, 2001).
2003	Weijters	Conjunto de técnicas y procedimientos que tratan de reconstruir de forma automática el modelo observado en los ficheros de registros (Weijters, 2003).
2009	Nakatumba & van der Aalst	Conjunto de técnicas que tienen como principal objetivo, descubrir, mejorar y monitorear los procesos reales a través de la extracción de conocimientos de un conjunto de ejecuciones reales de procesos (registro de eventos) (Nakatumba & van der Aalst, 2009).
2012	van der Aalst et al., 2012	Disciplina permite descubrir, monitorear y mejorar los procesos identificados en los sistemas de información que poseen las organizaciones a partir de sus registros de eventos (van der Aalst et al., 2012).
2014	Gupta	Disciplina que construye el puente entre la minería de datos como un enfoque de Business Intelligence y Business Process Management y tiene como objetivo el descubrimiento de modelos basados en el registro de eventos disponibles (Gupta, 2014).

2015	Aguirre	Disciplina que está compuesta por una serie de herramientas y técnicas basadas en la minería de datos para analizar procesos de negocio cuyo registro de eventos de ejecución real se encuentran disponibles en sistemas de información (Aguirre, 2015).
------	---------	--

El acervo referencial obtenido por el mapeo sistemático permitió identificar que la categoría 1) Fundamentos teóricos tuviera la mayor concentración de aportaciones conceptuales de manera general (Figura 3), y específicas tal como se visualiza en el eje y de la Figura 4. Por otro lado, en la categoría 2) Propuestas, se relaciona la integración de la minería de procesos con sistemas de información, modelos de referencia de procesos (Rubin, 2007; Samalikova, 2014; Valle et al., 2014b), modelos de evaluación de procesos (Gruaños, 2011; Valle et al., 2014), monitoreo y control de procesos.

La categoría 3) Sistemas de Información y herramientas agrupa las diversas técnicas y algoritmos de minería de procesos en herramientas de software comerciales: ARIS Process Performance Manager por Software AG, Disco por Fyxicon (Papadimitriou & Sun, 2008), Enterprise Visualization Suite por Businesscape, Interstage BPME por Fujitsu, Process Discovery Focus por Iontas, Reflect one por Pallas Athena y Reflect por Futura Process Intelligence. Por el contrario, la herramienta de minería de procesos de código abierto ProM es ampliamente utilizada en todo el mundo y proporciona un punto de partida sencillo para los profesionales, estudiantes y académicos (Van Der Aalst & Van Dongen, 2013).

En dichos sistemas de información se encontró que de acuerdo a la categoría 4) Algoritmos, se implementan principalmente tres algoritmos básicos: heurístico, difuso y genético (Gupta, 2014). La minería heurística proporciona una vista de flujos de trabajo científico, considerando la dependencia de larga distancia. Este algoritmo obtiene el modelo del proceso expresado en una C-net, es robusto ante el ruido y las ejecuciones incompletas de los procesos (Puldon & Llanes,

2005). El algoritmo difuso brinda una vista amplia del flujo de trabajo de forma científica mediante el control para mostrar una tarea diferente según el nivel de importancia. El algoritmo difuso devuelve el modelo de proceso en una red de Petri, con la limitante de no ser robusto ante el ruido, procesos poco estructurados o ejecuciones incompletas (Gupta, 2014). El tercer algoritmo trabaja con minería genética siendo capaz de obtener un modelo del proceso que refleja adecuadamente la realidad expresada en el registro de eventos pero tiene un alto costo computacional (Puldon & Llanes, 2005). El algoritmo genético proporciona una vista de la frecuencia en que se ejecutan las tareas y la sucesión entre las mismas. Así mismo, este algoritmo descubre todas las estructuras de control de flujo comunes (Gupta, 2014). El modelo extraído mediante cualquiera de los tres algoritmos de minería de procesos puede brindar una vista del(los) proceso(s) en un determinado punto de la ejecución; esto conduce a buscar diversas formas de analizar el proceso para la manipulación de los cambios en él.

En relación a las implicaciones de la minería de procesos con respecto a CMMI, Valle et al. (2014, 2014b) desarrolló un marco (estructura y contenido) para aplicar técnicas de minería de procesos en las evaluaciones SCAMPI con el objetivo de llevarla a cabo con una mayor profundidad y cobertura, mientras se mantiene el mismo nivel de esfuerzo. Los autores indican que la minería de procesos facilitaría el proceso de evaluación tipo SCAMPI, brindando la información que se requiere sobre las áreas de procesos definidas para un modelo de referencia de procesos, tal como CMMI-DEV v1.3.

El autor Samalikova (2014), realizó un estudio para identificar los elementos que apoyan las evaluaciones de procesos de software con CMMI. Identificó los elementos clave dentro de los objetivos y prácticas genéricas, encontrando en la minería de procesos una forma de apoyar los enfoques actuales para la evaluación y mejora de procesos de software. De igual manera, Gruaños (2011) investigó el potencial de la minería de procesos para apoyar un proyecto SPI y evaluación de procesos de software basado en CMMI. Propuso una metodología

concreta para encontrar las relaciones entre la minería de procesos y la evaluación de dos áreas de procesos de CMMI para un nivel 3 de madurez: Enfoque en Procesos de la Organización (OPF) y Análisis de Decisiones y Resolución (DAR). Concluyendo con que la minería de procesos es un camino viable para la recolección de datos en una evaluación y de esta forma reconstruir los procesos que se modelan con CMMI.

De esta forma, las aportaciones de (Samalikova, 2014; Valle et al., 2014, Valle et al., 2014b; Gruaños, 2011) recomiendan aprovechar, de forma significativa, los datos de eventos o registros de eventos en un proyecto SPI basado en CMMI-DEV permitiendo que los procesos de monitoreo y control de la ejecución de la implantación de las áreas de procesos, la gestión de proyectos, gestión de procesos y el de evaluación SCAMPI, por mencionar algunos se apoyarían de un modelo para la exploración y explotación del conocimiento asociado al cumplimiento de las áreas de procesos establecidos para un nivel específico de madurez y/o mejorar el modelo del proceso.

## 5. Conclusiones

Los hallazgos encontrados en el mapeo sistemático demuestran una evolución de la minería de procesos durante el periodo seleccionado, reconociéndose inicialmente como una técnica hasta llegar a formalizarse como una disciplina, con la inclusión de técnicas, algoritmos y herramientas basadas en principios de minería de datos, que permiten la exploración y explotación de información a partir de los registros de evento. Así mismo, se encontró evidencia entre la Minería de Procesos en proyectos SPI que han tenido como propósito apoyar los enfoques actuales de implementación y evaluación de procesos de software definidos por el modelo de referencia de procesos CMMI-DEV y su método de evaluación SCAMPI.

La evidencia encontrada en las cuatro categorías (Fundamentos teóricos, propuestas, herramientas y sistemas de información y algoritmos) representa

una base sobre la cual las organizaciones desarrolladoras de software encuentren una solución a los problemas que enfrentan asociados a la complejidad en el manejo de gran cantidad de información o a la falta de experiencia por parte del equipo de mejora en el uso de técnicas o algoritmos para ello. La implicación de la minería de procesos en este tipo de organizaciones es una alternativa para obtener una mayor visibilidad en la ejecución de las actividades requeridas por los modelos de procesos, a partir de la información almacenada en sus repositorios y concentrarse en aquellas que ayuden a satisfacer los objetivos del negocio permitiéndola ser más competitiva en la industria de software.

Actualmente, se está trabajando en un proyecto en el que se aplicarán los elementos conceptuales encontrados en esta revisión en una empresa mexicana certificada en CMMI-DEV con un el nivel 3 de madurez, para analizar los registros de eventos generados durante un año en las áreas de proceso de gestión de proyectos básicas, y de esta forma obtener un modelo del proceso que se compara con las evidencias requeridas en el proceso de evaluación SCAMPI.

## Agradecimientos

Este proyecto se encuentra registrado en la Coordinación de Posgrado e Investigación de la Universidad autónoma de Baja California por medio del programa 111/1298. Así mismo, se desea agradecer al CONACYT por el financiamiento otorgado al primer autor para realizar una estancia de investigación en la Universidad del Cauca (UNICAUCA).

# Referencias

Aguirre, S. (2015). Metodología para la aplicación de minería de procesos. Tesis Doctoral. Colombia.

Chrissis, M. B., Konrad, M., & Shrum, S. (2011). CMMI for Development®: Guidelines for Process Integration and Product Improvement. 3rd Edition, Addison-Wesley Professional.

CMMI Institute. (2015). Maturity Profile June 30 2015. <http://cmmiinstitute.com/resources/process-maturity-profile-july-2015>, Octubre 2015.

Coleman G. & O'Connor R. (2008). Investigating software process in practice: a grounded theory perspective. *Journal of Systems and Software*, 81, pp. 772–784.

Durón del Villar, B. L. & Muñoz-Mata, M. A. (2013). Selección de estrategias para la implementación de mejoras de procesos de Software. *ReCIBE*, 2, pp. 15.

Cruañas, J. R. (2010). "Process Mining Opportunities for CMMI Assessments" (Doctoral dissertation, Universitat Politècnica de Catalunya. Escola Tècnica Superior d'Enginyeria Industrial de Barcelona.

Flores Rios B.L., Astorga Vargas M.A, Rodríguez Elias O.M, Ibarra Esquer J.E. & Andrade M.C. (2014). Interpreting the mexican standards for software process implementation and capacity assessment under a knowledge management approach. *Revista de la Facultad de Ingeniería. Universidad de Antioquia*, 71, pp. 85-10.

Gupta, E. P. (2014). Process Mining a Comparative Study. *International Journal of Advanced Research in Computer and Communications Engineering*, 3, p. 5.

Kitchenham, B. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report, Software Engineering Group, Department of Computer Science, University of Durham.

Lester N. G., Wilkie F. G., McFall D., Ware M. P. (2010). Investigating the role of CMMI with expanding company size for small-to medium-sized enterprises. *Journal of Software Maintenance and Evolution: Research and Practice*, 22, pp. 17-31.

Martínez, Y., Castro, C. C., & Beigbeder, S. M. (2011). Evidencia empírica sobre mejoras en productividad y calidad en enfoques MDD: un mapeo sistemático. *REICIS: Revista Española de Innovación, Calidad e Ingeniería del Software*, 7(2), 6-27.



Nakatumba, J., & van der Aalst, W. M. (2009). Analyzing resource behavior using process mining. In International Conference on Business Process Management. Springer Berlin Heidelberg, pp. 69-80.

Niazi M. (2015). A comparative study of software process improvement implementation success factors. *Journal of Software: Evolution and Process*, 27 (9), pp. 700–722.

Niazi, M., Wilson, D. & Zowghi, D. (2005). A maturity model for the implementation of software process improvement: An empirical study. *Journal of Systems and Software*, 74(2), pp.155-172.

NYCE (2010). NMX-I-15504/03-NYCE-2010.Tecnología de la Información – Evaluación de los Procesos –Parte 03, Guía para realizar una evaluación. México.

Oktaba H. & Piattini M. (2008). Software Process Improvement for Small and Medium Enterprises: Techniques and Case Studies. IGI Global research collection.

Orellana García, A., Larrea Armenteros, O. U., & Pérez Alfonso, D. (2015). Generador de Registros de Eventos para el Análisis de procesos en el Sistema de Información Hospitalaria xavia HIS. Convención Salud. Disponible en <http://www.convencionsalud2015.sld.cu/index.php/convencionsalud/2015/paper/view/380/559>

Orellana García, A., Pérez Alfonso, D., & Estrada Sentí, V. (2016). Revisión de los principales modelos para aplicar técnicas de Minería de Procesos (Review of models for applying process mining techniques). *GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología*, 4(1).

Papadimitriou, S., & Sun, J. (2008). Disco: Distributed co-clustering with map-reduce: A case study towards petabyte-scale end-to-end mining. In 2008 Eighth IEEE International Conference on Data Mining (pp. 512-521). IEEE.

Pérez Alfonso, D., Yzquierdo Herrera, R., Silverio Castro<sup>1</sup>, R., & Lazo Cortes, M. (2012). Utilización de minería de proceso en el entorno empresarial cubano. 12 Ed.

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic mapping studies in software engineering. In 12th international conference on evaluation and assessment in software engineering, 17(1).

Pino, F., García, F., & Piattini, M. (2006). Revisión sistemática de mejora de procesos software en micro, pequeñas y medianas empresas. *Revista Española de Innovación, Calidad e Ingeniería del Software*, 2(1), pp. 6-23.

Pizarro Álvarez, G. E. (2013). Enfoque OLAP para el descubrimiento interactivo en minería de procesos. Tesis de Magíster en Ciencias de la Ingeniería. Pontificia Universidad Católica de Chile, 2013.

Puldon, J. J. & Llanes, R. S. (2005). Técnicas del algoritmo a y Minería Genética para la obtención de modelos de procesos a partir de trazas de ejecución de procesos. 7mo congreso de Iberoamericano de Ingeniería Mecánica.

Samalikova, J., Kusters, R. J., Trienekens, J. J. M., & Weijters, A. J. M. M. (2014). Process mining support for CMMI-based software process assessment, in principle and in practice. University of Technology Eindhoven, The Netherlands.

Secretaría de Economía (2016). Centros de Desarrollo Certificados/Verificados Vigentes en Modelos de Calidad. Disponible en [https://prosoft.economia.gob.mx/doc/PADRON\\_CENTRO%20DE%20DESARROLLO%20VIGENTE\\_2016\\_abr-18.pdf](https://prosoft.economia.gob.mx/doc/PADRON_CENTRO%20DE%20DESARROLLO%20VIGENTE_2016_abr-18.pdf)

Valle, A., Loures, E. R., & Portela, E. (2014). Process Mining Extension to SCAMPL. In SIMPDA (pp. 179-183).

Valle, A., Portela, E., Loures, E. R., & Deschamps, F. (2014b). Structure and Content of a Framework to Apply Process Mining Techniques in Scampi Assessments. In Proceedings of the International Annual Conference of the American Society for Engineering Management, p. 1.

Van der Aalst, W. M. P., (2011). Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Heidelberg.

Van der Aalst, W. M. P., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering, 16 (9), pp. 1128-1142.

Van der Aalst, W. M. P., Reijersa, A., Weijtersa, B., Van Dongena, A., Alves de Medeirosa M. S. & Verbeeka. (2007). Business process mining: An industrial application. Elsevier. 32(5), pp. 713–732.

Van der Aalst, W. M. P., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., et al.(2012). Process mining manifesto. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM 2011 Workshops, Part I. LNBIP, Springer, Heidelberg. pp. 169–194.

Van Der Aalst, W. M., & Van Dongen, B. F. (2013). Discovering petri nets from event logs. In Transactions on Petri Nets and Other Models of Concurrency VII, Springer Berlin Heidelberg. pp. 372-422.

Vasconcellos, F., Landre G., Cunha J.A., Oliveira J.L., Ferreira R. A. & Vincenzi Auri M.R. (2017). Approaches to strategic alignment of software process improvement: A systematic literature review. *The Journal of Systems and Software*, 123, pp. 45–63.

Verdonk M., van Werf J., van der Aalst, W. M. P., van Hee K. M. (2010). Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. *Computer*, 43(3), pp. 90-93.

Wang J., Wong R. K., Ding J., Guo Q. & Wen L. (2013). Efficient Selection of Process Mining Algorithms. *IEEE Transactions on Services Computing*, 6(4), p.13.

Weijters, A. J. M. M., (2003). Rediscovering Workflow Models from Event-Based Data using Little Thumb. *Integrated Computer-Aided Engineering*, 10(2), pp. 151-162.

Weijters, A. J. M. M., & van der Aalst, W. M. P. (2001). Process mining: discovering workflow models from event-based data. In *Belgium-Netherlands Conference on Artificial Intelligence*.

Yzquierdo Herrera, R. (2013). Minería de proceso como herramienta para la auditoria. *Ciencias de la Información*. 44(2), pp. 25-32.

## Notas biográficas:



**Paola Elvira Velazquez Solis** es Ingeniero en Computación de la UABC y actualmente, estudiante de Maestría en Ciencias en el Instituto de Ingeniería de la UABC. Su área de investigación es la Minería de Procesos. Se ha desempeñado como analista de sistemas, analista de procesos y consultor en el sector privado de México.



**Brenda Leticia Flores Rios** es Doctora en Ciencias por la UABC. Responsable del área de Ingeniería del conocimiento del Instituto de Ingeniería, desarrollando proyectos de investigación relacionados a la Gestión del Conocimiento y Mejora de procesos de software. Imparte docencia en licenciatura y posgrado. Pertenece a la Red

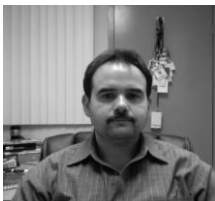
Temática Mexicana de Ingeniería de Software y a la Academia Mexicana de Computación. Ha participado como consultora en la implementación de la NMX-I-059-NYCE-2011 y como Miembro de Equipo Evaluador en SCAMPI A de CMMI-DEV nivel 3.



**María Angélica Astorga Vargas** es Maestra en Ciencias en el área de Computación. Profesora titular de tiempo completo en el Programa Educativo de Licenciado en Sistemas Computacionales de UABC. Su área de investigación es Modelos de Procesos de Software y Gestión del Conocimiento. Actualmente, se encuentra en formación de Doctorado. Pertenece a la Red Temática Mexicana de Ingeniería de Software. Ha participado como consultora en la implementación de la NMX-I-059-NYCE-2011 y como Miembro de Equipo Evaluador en SCAMPI A de CMMI-DEV nivel 3.



**Francisco J. Pino** es Doctor en Ingeniería Informática por la Universidad de Castilla-La Mancha (España). Ingeniero en Electrónica y Telecomunicaciones, y Especialista en Redes y Servicios Telemáticos de la Universidad del Cauca (Colombia). Profesor titular adscrito a la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca. Miembro del Grupo IDIS (Investigación y desarrollo en ingeniería de software) de la Universidad del Cauca. Investigador Senior reconocido por Colciencias. Consultor en calidad y mejora de procesos, productos y servicios de software. Auditor Jefe por AENOR de ISO 15504-SPICE. Sus intereses de investigación y profesionales se enfocan en a área de calidad y mejora de procesos de desarrollo de software en pequeñas empresas y en entornos multimodelos. Sobre estos temas ha sido autor de varios libros y diversos artículos en revistas y congresos nacionales e internacionales.



**Jorge Eduardo Ibarra Esquer** es Maestro en Ciencias por el Centro de Investigación Científica y de Educación Superior de Ensenada. Se desempeñó como coordinador del Programa Educativo de Ingeniero en Computación de UABC. Actualmente, es profesor titular de tiempo completo y desarrolla líneas de investigación en enseñanza de la programación, sistemas electrónicos digitales e internet de las cosas. Pertenece a la Red Temática Mexicana de Ingeniería de Software.



**Félix Fernando González-Navarro** es Doctor en Inteligencia Artificial por el Departamento de Lenguajes y Sistemas Informáticos de la Universitat Politècnica de Catalunya. Actualmente, es investigador titular de tiempo completo en la UABC liderando el Laboratorio de Inteligencia y líder del Cuerpo Académico de Cómputo Científico ante PRODEP-SEP. Pertenece a la Red Temática Mexicana de Ingeniería de Software y es miembro del Sistema Nacional de Investigadores (SNI) nivel 1.



Esta obra está bajo una licencia de Creative Commons  
Reconocimiento-NoComercial-CompartirIgual 2.5 México.